

Sarah F. Majors

(724) 683 - 5500
Leetsdale, Pennsylvania

sfmajors373@gmail.com
www.github.com/sfmajors373

MACHINE LEARNING PROJECTS

Breast Cancer Capstone

- Neural network was designed and trained against stained lymph node slides, from the Camelyon 16 dataset, to perform image segmentation with the goal of the detection of cancerous and abnormal cell structures
- Various methods such as OTSU were used to efficiently clean, tile, and mask the slide data to reduce man hours necessary for production of a dataset as well as enhance the quality of the training set and model accuracy
- Convolutional Neural Network model was designed to flag images from Camelyon 16 dataset as abnormal
- UNet based image segmentation was performed against Camelyon 16 dataset images identified by the CNN
- Pix2Pix Generative Adversarial Network was used to expand the existing Camelyon 16 dataset to create more training data to feed the pipeline to enhance reliability and accuracy by balancing positive and negative samples
- Both GPU and CPU computation were utilized to reduce training time and increase the velocity of development
- Set up and deployed a Jupyter Labs based research environment in a Docker container based on Arch Linux on a 32 core Threadripper machine with a mirror of the Camelyon 16 dataset for each team member to reduce the set up time and allow for rapid iterations
- Deployed the model in a flask application on Google Cloud allowing a user to upload a whole slide or just an image tile and see the prediction as well as the segmentation, if applicable

Privacy Preserving Satellite Imagery

- Created reusable Docker container housing Jupyter Labs and all necessary machine learning libraries to reduce issues related to dependency management
- Compiled PySyft and PyGrid packages for Arch Linux and became the maintainer of those and several other machine learning packages in the Arch User Repository
- Built computer to reduce dependency on expensive cloud compute and prevent runaway cost while experimenting with various models, parameters, etc as well as speed up training time
- Worked on a demo intended for the United Nations that would utilize privacy preserving methods to pull data from satellite imagery without the end user ever actually seeing the images

CheXpert

- Utilized the CheXpert dataset to develop a potential low cost solution to detect chest pathologies deployed to Raspberry Pi via Flask with three other people
- Using VGG16 we obtained a 79.15% accuracy rate with a 50,000 image training set

Federated Learning Cluster

- Created a federated learning cluster on a set of four Raspberry Pi 4s to have a platform to further explore how federated learning works
- Practiced compiling Arch packages for ARM so I could use Arch on the Pis

Noise Analysis on Android

- Worked with three others to compare Gaussian and Laplacian noise as the Laplace mechanism in differential privacy applications in an Android context

CONTINUING EDUCATION

Machine Learning Fundamentals

In Progress

- Completed the AI in Medicine Specialization by deeplearning.ai
- Competing in Intel Edge AI scholarship challenge
- Competing in Bertelsman AI Deep Learning scholarship challenge
- Completed the Computer Vision Nano Degree from Udacity focusing on applying computer vision methods with the PyTorch and OpenCV libraries, projects include facial keypoint detection, image captioning and SLAM
- Completed and won the Secure and Private AI Facebook Scholarship Challenge Course focusing on differential and federated learning using PySyft and PyTorch with a heavy emphasis on group projects coordinated in the Slack community
- Completed and received certificate for the Deep Learning Specialization from deeplearning.ai covering deep neural networks, convolutional neural networks, sequence models and structuring machine learning projects
- Completed Machine Learning course by deeplearning.ai and Andrew Ng using Octave
- Completed the Bertelsman Data Science Challenge Scholarship Course focusing on statistics, python, and SQL
- Working to enhance my math understanding through MITx and Outlier

EDUCATION

Fourth Brain

2020

Machine Learning

LambdaSchool

2017-2018

- Completed the full time 6 month computer science course through Lambda School
- Learned data structures, algorithms, front end, back end, middleware, and databases
- Tech stack: JavaScript, React, Express, Node.js, MongoDB, Docker, Kubernetes
- Completed an additional 5 week full time course specializing in Java backend
- Tech stack: Java, Spring, Maven, RabbitMQ, SQLite, Swagger

Mercyhurst University

Bachelors of Science in Anthropology

WORK HISTORY

Teamsense

January 2022 - June 2022

- Code, bugs, implementation design contribution, support person
- Hardware stuff

Rivers Agile

April 2019 - January 2022

- Converted existing code to VueJS in order to add internationalization features to prepare for use in Europe
- Chased bugs, created features, and developed prototypes for real time dashboards and performance reports to showcase the efficiency of the autonomous forklifts to the clients with the goal of aiding in sales
- Participated in planning meetings for new software projects in which there was integration with other teams in order to fully understand customer needs from the product and then created user stories from that perspective
- Participated in architectural planning meetings for new projects and tested different technologies by creating rapid prototypes and discussing the pros and cons of different approaches
- Tech Stack: Python, Plot.ly, Django, RabbitMQ, MQTT, VueJS, node, selenium, i18n, Docker, Vagrant

Contractor

February 2018 - January 2019

- Collected data from the FAA pertaining to yearly flight hours per aircraft model and from the NTSB pertaining to aircraft accident data by creating working relationships with officials from both organizations
- Ran statistical analyses on data to define a standard metric to compare the safety of different aircraft types, determined accidents per 100,000 hours of flight was most accurate and relevant; calculated this for overall safety as well as in differing environmental conditions
- Created and deployed a website in Golang to display the metrics, which were graphed using Matplotlib
- Recorded detailed methods regarding the statistical analysis to ensure reproducibility and accountability
- Created a bot to scrape the NTSB database and alert subscribed twitter users to when a final report was released
- Tech stack: Python, Jupyter Notebooks, Golang, SQL, Matplotlib

Nightingale Security

October 2017 - February 2018

- Used the bug tracker to locate and correct various bugs in the drone control UI, such as display and logic errors
- Created last minute fixes for UX issues in time for product demonstrations to clients and venture capitalists
- Designed a feature to allow the user to create bounding boxes on images which were to be fed into a machine learning algorithm which would then be utilized by the drone to automatically identify risks
- Utilized agile development methods in a remote setting, including Kanban, Slack, and video conferencing
- Tech stack: JavaScript, PHP, Angular, GoogleMapsAPI, Git

Topographical Data Analysis

January - April 2016

- Reverse engineered the recording of data from the Tripod Data Systems Survey Pro in order to correct errors in manual entry of information into the total station and prevent the loss of a day's measurements
- Utilized the data from the total station to make accurate 3D images of the topography and identify systemic issues in our mapping policies including spacing and user carelessness
- Tech stack: QGIS, ROOT (CERN), Python, LaTeX